Foundation of Cryptography, Lecture 4
Pseudorandom Functions

Handout Mode

Iftach Haitner, Tel Aviv University

Tel Aviv University.

March 11, 2014

Motivation Discussion

1. We've seen a small set of objects: $\{G(x)\}_{x \in \{0,1\}^n}$, that "looks like" a larger set of objects: $\{x\}_{x \in \{0,1\}^{2n}}$.
2. We want small set of objects: *efficient function families*, that looks like a huge set of objects: *the set of all functions*.

Solution

Function families

1. $\mathcal{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}}$, where $\mathcal{F}_n = \{f\colon \{0,1\}^{m(n)} \mapsto \{0,1\}^{\ell(n)}\}$
2. We write $\mathcal{F} = \{\mathcal{F}_n\colon \{0,1\}^{m(n)} \mapsto \{0,1\}^{\ell(n)}\}$
3. If $m(n) = \ell(n) = n$, we omit it from the notation
4. We identify function with their description

Random functions

Definition 1 (random functions)

For $n, k \in \mathbb{N}$, let $\Pi_{n,k}$ be the family of all functions from $\{0,1\}^n$ to $\{0,1\}^k$.
Let $\Pi_n = \Pi_{n,n}$.

- $\pi \stackrel{R}{\leftarrow} \Pi_n$ is a “random access" source of randomness
- Parties with access to a common $\pi \stackrel{R}{\leftarrow} \Pi_n$ can do a lot

- How long does it take to describe $\pi \in \Pi_n$? $2^n \cdot n$ bits
- The truth table of $\pi \stackrel{R}{\leftarrow} \Pi_n$ is a uniform string of length $2^n \cdot n$
- For integer function m, we will consider the function family $\{\Pi_{n,m(n)}\}$.

Efficient function families

Definition 2 (efficient function family)

An ensemble of function families $\mathcal{F} = \{\mathcal{F}_n\}_{n\in\mathbb{N}}$ is efficient, if:

Samplable. $\mathcal{F}$ is samplable in polynomial-time: there exists a PPT that given 1^n, outputs (the description of) a uniform element in $\mathcal{F}_n$.

Efficient. There exists a polynomial-time algorithm that given $x \in \{0,1\}^n$ and (a description of) $f \in \mathcal{F}_n$, outputs $f(x)$.

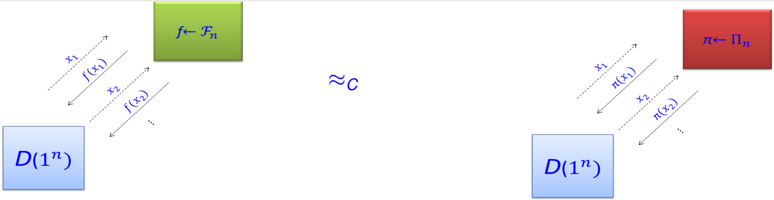
Pseudorandom Functions

Definition 3 (pseudorandom functions (PRFs))

An efficient function family ensemble $\mathcal{F} = \{\mathcal{F}_n \colon \{0,1\}^{m(n)} \mapsto \{0,1\}^{\ell(n)}\}$ is pseudorandom, if

$$\left| \Pr_{f \xleftarrow{R} \mathcal{F}_n} \left[\mathsf{D}^f(1^n) = 1 \right] - \Pr_{\pi \xleftarrow{R} \Pi_{m(n),\ell(n)}} \left[\mathsf{D}^\pi(1^n) = 1 \right] \right| = \operatorname{neg}(n),$$

for any oracle-aided PPT D.



- Why "oracle-aided"?
- Easy to construct (no assumption!) with logarithmic input length
- PRFs of super logarithmic input length, which is the interesting case, imply PRGs
- We will mainly focus on the case $m(n) = \ell(n) = n$
- We write $\mathsf{D}^{\mathcal{F}}$ to stand for $(\mathsf{D}^f)_{f \xleftarrow{R} \mathcal{F}}$.

Section 2

PRF from OWF

Naive Construction

Let $G\colon \{0,1\}^n \mapsto \{0,1\}^{2n}$, and for $s \in \{0,1\}^n$ define $f_s\colon \{0,1\} \mapsto \{0,1\}^n$ by

- $f_s(0) = G(s)_{1,\ldots,n}$
- $f_s(1) = G(s)_{n_1,\ldots,2n}$.

Claim 4

Assume G is a PRG, then $\{\mathcal{F}_n = \{f_s\}_{s\in\{0,1\}^n}\}_{n\in\mathbb{N}}$ is a PRF.

Proof: The truth table of $f \stackrel{R}{\leftarrow} \mathcal{F}_n$ is $G(U_n)$, where the truth table of $\pi \stackrel{R}{\leftarrow} \Pi_{1,n}$ is U_{2n}□

- Naturally extends to input of length $O(\log n)$:-)
- Miserably fails for longer length (which is the only interesting case) :-(
- Problem, we are constructing the whole truth table, even to compute a single output

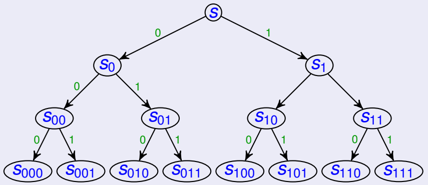
The GGM Construction

Construction 5 (GGM)

For $G\colon \{0,1\}^n \mapsto \{0,1\}^{2n}$ and $s \in \{0,1\}^n$,

- $G_0(s) = G(s)_{1,\ldots,n}$
- $G_1(s) = G(s)_{n+1,\ldots,2n}$

For $x \in \{0,1\}^k$ let $f_s(x) = G_{x_k}(f_s(x_{1,\ldots,k-1}))$, letting $f_s() = s$.



$s_x = f_s(x)$

- Example: $f_s(001) = s_{001} = G_1(s_{00}) = G_1(G_0(s_0)) = G_1(G_0(G_0(s)))$
- G is poly-time $\implies \mathcal{F} := \{\mathcal{F}_n = \{f_s\colon s \in \{0,1\}^n\}\}$ is efficient

Theorem 6 (Goldreich-Goldwasser-Micali (GGM))

If G is a PRG then $\mathcal{F}$ is a PRF.

Corollary 7

OWFs imply PRFs.

Proof Idea

Assume $\exists$ PPT D, $p \in$ poly and infinite set $\mathcal{I} \subseteq \mathbb{N}$ with

$$\left|\Pr[\mathsf{D}^{\mathcal{F}_n}(1^n) = 1] - \Pr[\mathsf{D}^{\Pi_n}(1^n) = 1]\right| \geq \frac{1}{p(n)}, \tag{1}$$

for any $n \in \mathcal{I}$.

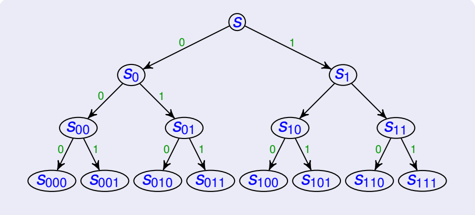
Fix $n \in \mathbb{N}$ and let $t = t(n)$ be a bound on the running time of $\mathsf{D}(1^n)$. We use D to construct a PPT D′ such that

$$\left|\Pr[\mathsf{D}'((U_{2n})^t) = 1] - \Pr[\mathsf{D}'(G(U_n))^t) = 1\right| > \frac{1}{np(n)},$$

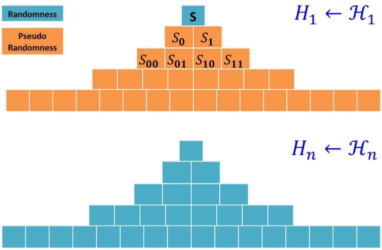
where $(U_{2n})^t = U_{2n}^{(1)}, \ldots, U_{2n}^{(t)}$ and $G(U_n)^t = G(U_n^{(1)}), \ldots, G(U_n^{(t)})$.

Hence, D′ violates the security of G.(?)

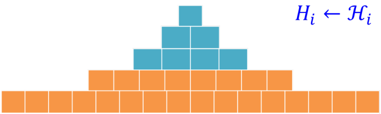
The Hybrid



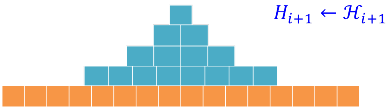
$s_x = f_s(x)$



- Let $\mathcal{T}_i$ be the set of all possible trees, in which the $i+1,\ldots,n$ levels are obtained by “applying GGM" to the i'th level.
- Given a tree t, let $h_t(x)$ return the x'th leaf of t.
- What family is $\mathcal{H}_1 = \{h_t\}_{t \in \mathcal{T}_1}$? $\mathcal{F}_n$. What is $\mathcal{H}_n$? Π_n.
- For some $i \in \{1,\ldots,i-1\}$, algorithm D distinguishes $\mathcal{H}_i$ from $\mathcal{H}_{i+1}$ by $\frac{1}{np(n)}$

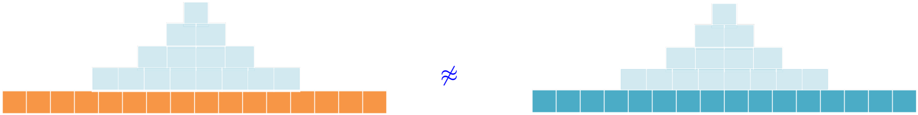


$\not\approx$



The Hybrid cont.

We assume wlg. that D distinguishes between $\mathcal{H}_{n-1}$ and $\mathcal{H}_n$ (?)

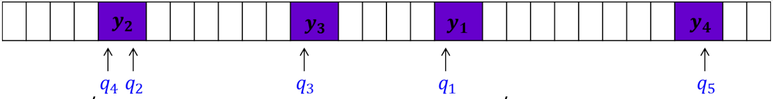


- D distinguishes (via t samples) between
  - R – a uniform string of length $2^n \cdot n$, and
  - P - a string generated by 2^{n-1} independent calls to G
- We would like to use D for breaking the security of G, but R and P seem too long :-(
- Solution: focus on the part (i.e., cells) that D *sees*

Algorithm 8 (D′ on $y_1, \ldots, y_t \in (\{0,1\}^{2n})^t$)

Emulate D. On the i'th query q_i made by D:

- If the cell queries by q_i'th is empty, fill it with the next y
- Answer with the content of the q_i'th cell.



- $\mathrm{D}'((U_{2n})^t)$ / $\mathrm{D}'(G(U_n))^t)$ emulates D with access to R / P
- Hence, $\left|\Pr[\mathrm{D}'((U_{2n})^t) = 1] - \Pr[\mathrm{D}'(G(U_n))^t) = 1\right| > \frac{1}{np(n)}$

Part I

Pseudorandom Permutations

Formal Definition

Let $\widetilde{\Pi}_n$ be the set of all permutations over $\{0,1\}^n$.

Definition 9 (pseudorandom permutations (PRPs))

A *permutation* ensemble $\mathcal{F} = \{\mathcal{F}_n : \{0,1\}^n \mapsto \{0,1\}^n\}$ is a pseudorandom permutation, if

$$\left|\Pr[\mathsf{D}^{\mathcal{F}_n}(1^n) = 1] - \Pr[\mathsf{D}^{\widetilde{\Pi}_n}(1^n) = 1\right| = \operatorname{neg}(n), \tag{2}$$

for any oracle-aided PPT D

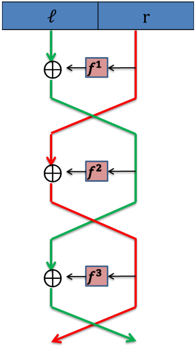
- Eq 2 holds for any PRF (taking the role of $\mathcal{F}$)
- Hence, PRPs are indistinguishable from PRFs...
- If no one can distinguish between PRFs and PRPs, let's use PRFs
  - (partial) Perfect “security"
  - Inversion

Section 3

PRP from PRF

Feistel Permutation

How does one turn a function into a permutation?



Definition 10 (LR)

For $f \colon \{0,1\}^n \mapsto \{0,1\}^n$, let $\mathrm{LR}_f \colon \{0,1\}^{2n} \mapsto \{0,1\}^{2n}$ be defined by

$$\mathrm{LR}_f(\ell, r) = (r, f(r) \oplus \ell).$$

- LR_f is a permutation: $\mathrm{LR}_f^{-1}(z, w) = (f(z) \oplus w, z)$
- LR_f is efficiently computable and invertible given oracle access to f
- For $i \in \mathbb{N}$ and $f^1, \ldots, f^i$, define $\mathrm{LR}_{f^1,\ldots,f^i} \colon \{0,1\}^{2n} \mapsto \{0,1\}^{2n}$ by

  $\mathrm{LR}_{f^1,\ldots,f^i}(\ell, r) = (r^{i-1}, f^i(r^{i-1}) \oplus \ell^{i-1})$, for $(\ell^{i-1}, r^{i-1}) = \mathrm{LR}_{f^1,\ldots,f^{i-1}}(\ell, r)$.
  (letting $(\ell^0, r^0) = (\ell, r)$)

Luby-Rackoff Thm.

Recall $\mathsf{LR}_f(\ell, r) = (r, f(r) \oplus \ell)$.

Definition 11

Given a function family $\mathcal{F} = \{\mathcal{F}_n \colon \{0,1\}^n \mapsto \{0,1\}^n\}$, let $\mathsf{LR}^i(\mathcal{F}) = \{\mathsf{LR}^i_{\mathcal{F}_n} = \{\mathsf{LR}_{f^1,\ldots,f^i} \colon f^1, \ldots, f^i \in \mathcal{F}_n\}\}$,

- $\mathsf{LR}^i_{\mathcal{F}}$ is always a permutation family, and is efficient if $\mathcal{F}$ is.
- Is $\mathsf{LR}^1_{\mathcal{F}}$ pseudorandom?
- $\mathsf{LR}^2_{\mathcal{F}}$? $\mathsf{LR}_{f^1,f^2}(0^n, 0^n) = \mathsf{LR}_{f^2}(0^n, f^1(0^n)) = (f^1(0^n), \cdot)$
  and $\mathsf{LR}_{f^1,f^2}(1^n, 0^n) = \mathsf{LR}_{f^2}(0^n, f^1(0^n) \oplus 1^n) = (f^1(0^n) \oplus 1^n, \cdot)$
- $\mathsf{LR}^3_{\mathcal{F}}$?

Theorem 12 (Luby-Rackoff)

Assuming that $\mathcal{F}$ is a PRF, then $\mathsf{LR}^3_{\mathcal{F}}$ is a PRP

- $\mathsf{LR}^4(\mathcal{F})$ is pseudorandom even if inversion queries are allowed

Proving Luby-Rackoff

It suffices to prove that $\mathrm{LR}^3_{\Pi_n}$ is pseudorandom (?)

- How would you prove that?
- Maybe $\mathrm{LR}^3(\Pi_n) \equiv \widetilde{\Pi}_{2n}$? description length of element in $\mathrm{LR}^3(\Pi_n)$ is $2^n \cdot 3n$, where that of element in $\widetilde{\Pi}_{2n}$ is $\log(2^{2n}!) > \log\left((\frac{2^{2n}}{e})^{2^{2n}}\right) > 2^{2n} \cdot n$

Claim 13

For any q-query D,

$$|\Pr[\mathrm{D}^{\mathrm{LR}^3(\Pi_n)}(1^n) = 1] - \Pr[\mathrm{D}^{\widetilde{\Pi}_{2n}}(1^n)| = 1] \in O(q^2/2^n).$$

- We assume for simplicity that D is *deterministic*, *non-repeating* and *non-adaptive*.
- Let $x_0, x_1, \ldots, x_q$ be D's queries.
- We show $(f(x_0), \ldots, f(x_q))_{f \xleftarrow{R} \mathrm{LR}^3(\Pi_n)}$ is $O(q^2/2^n)$ close (i.e., in statistical distance) to $(f(x_0), \ldots, f(x_q))_{f \xleftarrow{R} \widetilde{\Pi}}$
- To do that, we show both distributions are $O(q^2/2^n)$ close to $\mathit{Distinct} := \left((z_1, \ldots z_q) \xleftarrow{R} (\{0,1\}^{2n})^q \mid \forall i \neq j \colon (z_i)_0 \neq (z_j)_0\right)$.

Reminder: Statistical Distance

Definition 14

The statistical distance between distributions P and Q over $\mathcal{U}$, is defined by

$$\mathsf{SD}(P,Q) = \frac{1}{2} \cdot \sum_{u \in \mathcal{U}} |P(u) - Q(u)| = \max_{\mathcal{S} \subseteq \mathcal{U}} \{\Pr_Q[\mathcal{S}] - \Pr_P[\mathcal{S}]\}$$

In case $\mathsf{SD}(P,Q) \leq \varepsilon$, we say that P and Q are ε close.

Fact 15

Let $\mathcal{E}$ be an event (i.e., set) and assume $\mathsf{SD}(P|_{\neg\mathcal{E}}, Q) \leq \delta_1$ and $\Pr_P[\mathcal{E}] \leq \delta_2$. Then $\mathsf{SD}(P,Q) \leq \delta_1 + \delta_2$

Proving Fact 15

For any set $\mathcal{S}$, it holds that

$$\begin{aligned}\Pr_P[\mathcal{S}] &= \Pr_P[\mathcal{E}] \cdot \Pr_{P|_{\mathcal{E}}}[\mathcal{S}] + \Pr_P[\neg\mathcal{E}] \cdot \Pr_{P|_{\neg\mathcal{E}}}[\mathcal{S}] \\ &\geq (1-\delta_2) \cdot \Pr_{P|_{\neg\mathcal{E}}}[\mathcal{S}]\end{aligned} \tag{3}$$

Hence,

$$\begin{aligned}\Pr_Q[\mathcal{S}] - \Pr_P[\mathcal{S}] &\leq \Pr_Q[\mathcal{S}] - (1-\delta_2) \Pr_{P|_{\neg\mathcal{E}}}[\mathcal{S}] \\ &\leq \Pr_Q[\mathcal{S}] - \Pr_{P|_{\neg\mathcal{E}}}[\mathcal{S}] + \delta_2\end{aligned} \tag{4}$$

Thus,

$$\mathrm{SD}(P,Q) = \max_{\mathcal{S}}\{\Pr_Q[\mathcal{S}] - \Pr_P[\mathcal{S}]\} \leq \max_{\mathcal{S}}\{\Pr_Q[\mathcal{S}] - \Pr_{P|_{\neg\mathcal{E}}}[\mathcal{S}]\} + \delta_2 = \delta_1 + \delta_2.$$

$(f(x_0),\ldots,f(x_q))_{f \xleftarrow{R} \widetilde{\Pi}}$ is close to *Distinct*

Recall *Distinct* $:= \left((z_1,\ldots z_q) \xleftarrow{R} (\{0,1\}^{2n})^q \mid \forall i \neq j \colon (z_i)_0 \neq (z_j)_0\right)$.

For $f \in \widetilde{\Pi}$, let $Bad(f) := \exists i \neq j \colon f(x_i)_0 = f(x_j)_0$.

Claim 16

$\Pr_{f \xleftarrow{R} \widetilde{\Pi}}[Bad(f)] \leq \frac{\binom{q}{2}}{2^n} \leq \frac{q^2}{2^n}$

Proof: ?

Claim 17

$\left((f(x_0),\ldots,f(x_q)); f \xleftarrow{R} \widetilde{\Pi} \mid \neg \mathrm{Bad}(f)\right) \equiv \mathit{Distinct}$

Proof: ?

By Fact 15, $(f(x_0),\ldots,f(x_q))_{f \xleftarrow{R} \widetilde{\Pi}}$ is $\frac{q^2}{2^n}$ close to *Distinct*

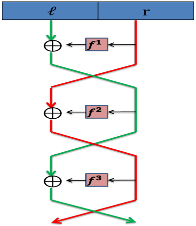
$(f(x_0),\dots,f(x_q))_{f \xleftarrow{R} \mathrm{LR}^3(\Pi_n)}$ is close to *Distinct*

Let $(\ell_1^0, r_1^0),\dots,(\ell_q^0, r_q^0) = (x_1,\dots,x_k)$.

The following rv's are defined w.r.t. $(f^1, f^2, f^3) \xleftarrow{R} \Pi_n^3$.

ℓ_1^0	r_1^0	ℓ_2^0	r_2^0	$\dots$	ℓ_q^0	r_q^0
ℓ_1^1	r_1^1	ℓ_2^1	r_2^1	$\dots$	ℓ_q^1	r_q^1
ℓ_1^2	r_1^2	ℓ_2^2	r_2^0	$\dots$	ℓ_q^2	r_q^2
ℓ_1^3	r_1^3	ℓ_2^3	r_2^0	$\dots$	ℓ_q^3	r_q^3

where $\ell_b^j = r_b^{j-1}$ and $r_b^j = f^j(r_b^{j-1}) \oplus \ell_b^{j-1}$.



Claim 18

$$\Pr_{f^1 \xleftarrow{R} \Pi_n}\left[\mathrm{Bad}^1 := \exists i \neq j \colon r_i^1 = r_j^1\right] \leq \frac{\binom{q}{2}}{2^n}$$

Proof: $r_i^0 = r_j^0 \implies r_i^1 \neq r_j^1$ and $r_i^0 \neq r_j^0 \implies \Pr_{f^1}\left[r_i^1 = r_j^1\right] = 2^{-n}$ $\square$

Claim 19

$$\Pr_{(f^1,f^2) \xleftarrow{R} \Pi_n^2}\left[\mathrm{Bad}^2 := \exists i \neq j \colon r_i^1 = r_j^1 \vee r_i^2 = r_j^2\right] \leq 2 \cdot \frac{\binom{q}{2}}{2^n} \in O\left(\frac{q^2}{2^n}\right)$$

Proof: similar to the above

Claim 20

$$\left((\ell_1^3, r_1^3),\dots,(\ell_q^3, r_q^3) \mid \neg\mathrm{Bad}^2\right) \equiv \mathit{Distinct}$$

Proof: ?

Proving Claim 20

Let $\mathcal{S} = \{(z_1, \ldots, z_q) \in (\{0,1\}^n)^q \colon \forall i \neq j \colon z_i \neq z_j\}$.

Claim 21

$\left((\ell_1^3, \ldots, \ell_q^3) \mid \neg \mathsf{Bad}^2\right)$ is uniform over $\mathcal{S}$.

Proof: For any $\mathbf{z} = (z_1, \ldots, z_q) \in (\{0,1\}^n)^q$ and $\pi \in \Pi_n$:

$$\Pr\left[(\ell_1^3, \ldots, \ell_q^3) = \mathbf{z}\right] = \Pr\left[(\ell_1^3, \ldots, \ell_q^3) = \pi(\mathbf{z}) := (\pi(z_1), \ldots, \pi(z_q))\right] \square$$

Section 4

Applications

General paradigm

Design a scheme assuming that you have random functions, and the realize them using PRFs.

Private-key Encryption

Construction 22 (PRF-based encryption)

Given an (efficient) PRF $\mathcal{F}$, define the encryption scheme $(\mathrm{Gen}, \mathrm{E}, \mathrm{D})$):

Key generation: $\mathrm{Gen}(1^n)$ returns $k \stackrel{\mathrm{R}}{\leftarrow} \mathcal{F}_n$

Encryption: $\mathrm{E}_k(m)$ returns $U_n, k(U_n) \oplus m$

Decryption: $\mathrm{D}_k(c = (c_1, c_n))$ returns $k(c_1) \oplus c_2$

- Advantages over the PRG based scheme?
- Proof of security?

Conclusion

- We constructed PRFs and PRPs from length-doubling PRG (and thus from one-way functions)
- Main question: find a simpler, more efficient construction

  or at least, a less adaptive one